数据新闻产生过程中的新闻意识

摘 要：随着互联网技术的发展，大数据在各行各业显现出重要的价值。在新闻行业，通过大数据工具对新闻要素进行挖掘、分析，形成了不同于传统新闻媒体的一种报道方式，即数据新闻。数据新闻在生产的过程中依赖互联网大数据。本文通过对数据新闻产生过程的剖析，厘清数据新闻产生过程中各个环节的作用机制，有利于数据新闻更好地发挥其特色，为受众报道事实。
关键词：数据新闻；大数据；生产
中图分类号：G210.7 文献标识码：A
文章编号：1671-0134（2018）06-105-02 DOI：10.19483/j.cnki.11-4653/n.2018.06.039

文 / 李璇 王明明

1. 数据新闻

数据新闻是大数据对新闻行业渗透的产物。数据新闻具有两个最显著的特点：（1）数据的支撑和直观的呈现方式。数据的支持可以细分到两点，一是数据新闻的价值衡量可以通过对数据库里的数据进行抓取、分析，依据受众的信息需求欲望给予一定的价值判断；二是数据新闻报道方式的内容是充满着数据的，是用数据呈现新闻报道的主题和角度的，是通过对数据相关性的分析来报道事实。（2）数据新闻的呈现方式是可视化的，是直观简洁的，这是它对传统的新闻报道方式最为明显的改进。

2. 数据新闻产生的过程

2.1 选题

选题是数据新闻的第一步。做数据新闻的选题要注重两方面因素：一是新闻价值，二是这个内容是否适用于数据新闻的呈现方式。

数据的收集和整理需要耗费一定的时间，因此，在时效性方面可能比不上传统新闻。在选题策划的时候要充分考虑新闻类型对时效性的要求。数据新闻揭示的是数据背后的逻辑关系，往往是对现有的新闻数据的再次使用和深挖，发现其背后的关系，这无疑决定数据新闻在确定选题时必然注重选题显著性，这样才使新闻数据有挖掘的意义。选题的接近性是指选题与受众的关联意识。在策划选题前，选题与受众的接近性，这种接近性的程度有多高？这都是应该考虑的，是通过大数据可以得到这种关联程度的准确排位的。

数据新闻的报道特点是将数据间的相互关系通过可视化的技术呈现在观众面前，这里的呈现类型包括图片和文字、音频、视频等，都是数据新闻可视化呈现方式。数据新闻主要适用于一些数据量大且数据之间存在某种关系，这种关系的揭示和深挖具有新闻价值。

2.2 数据的收集和筛选

数据新闻数据的收集因为数据的广泛性需要根据选题有针对性地收集数据。数据新闻数据的收集主要是通过爬虫工具实现，比如 Arachnid、Crawlzilla、Python 等数据的筛选，是确保新闻真实性的有效途径。在数据收集的同时进行数据筛选可以为下一步数据分析提供一个有价值的分析量。数据的真实性取决于数据来源的可信度。目前，数据新闻报道中的数据大都来源于政府公开的数据、非政府机构收集的数据、企业等。在数据收集之后进行筛选，主要是对收集的数据进行清洗。非政府机构收集的数据是否全面真实？企业提供的数据是否出于利益有所失实？媒体自我收集的数据是否带有片面性？这些都是在数据清洗过程中要注意的，清洗掉那些不真实、不必要的数据，为下一步数据分析提供一个良好的环境。

2.3 数据的分析

数据分析可以从数据中发现数据背后的故事和价值，让数据能够成功地成为新闻。数据的分析还可以挖掘出好的新闻选题。

面对杂乱无章的数据，分析者首先要将收集到的数据统一化，便于观察和分析；其次，在数据分析时要注重每一个数据的关系变化。数据分析可以明确数据新闻的报道内容，不同的分析层面决定了数据新闻不同的深度和广度，以及报道的故事性呈现方式。在数据分析的过程中，关注数据极值的出现，注重数据之间的对比关系和逻辑关系，善于观察数据发展的趋势，大胆地推测并进行有效的验证，这些都是数据新闻分析中常用的方法。数据分析的工具主要有：ERP、Matlab、SAS、SPSS 等，这些工具都是较专业的数据分析工具。

2.4 数据新闻的呈现

数据新闻的可视化呈现形式主要包括各种静动态图表。根据具体的主题选用不同的背景类型的图表形式，从而将数据新闻信息的逻辑关系秩序化和形象化。目前出现的图表形式主要有以地图为背景的、以罗盘为背景的，以地平线为背景的，以城市为背景的和山脉为背景的等，还有一些流程图、树状图、热力图和散点图等，具体哪种类型比较适合所需报道的数据新闻内容，这要结合具体的数据意义来选择。目前广泛使用的可视化制作工具主要包括 icharts 和 tableau 等。最后，一定要对可视化的图表进行润色，比如对图标中的每一类数据进项

颜色区分，帮助受众更直观地厘清信息之间的区别。

数据新闻的产生基于一个团队，这个团队中必须有四种角色，即记者编辑、数据分析师、美术设计师和程序员。内容的组织由记者编辑负责，数据的搜集、筛选由数据分析师负责，图形的设计与优化由美术设计师负责，程序员主要负责由代码呈现图形。而在实际操作中，往往一个人会身兼数职。

3. 产生过程中存在的问题

3.1 数据的质量

数据新闻中数据的获取是通过政府机构获得共享数据，或从非政府机构和企业那里获得数据，很少有新闻媒体是自己搜集数据的，一是工程量大，二是数据获取渠道的闭塞和有限。通过从他人那里获得数据毕竟是有限的，而数据新闻又是由数据驱动的，是离不开数据的，因此，数据的获取成为了数据新闻发展的一个障碍。

现下的许多数据新闻作品中都出现大量的数据，专业的做法是对数据的来源进行标注，但仍不乏一些新闻媒体对数据来源不予交代，因为这些数据可能来源于别的媒体报道之中，也就是来源于第三者之手，第三者一旦不具有权威性，就会使得新闻的说服力不强。这样做一方面会对媒体的形象造成损失，降低受众对这一媒体的信任度；另一方面，数据新闻的逻辑分析经由错误的数据得到加深和呈现，很有可能会造成假新闻在受众之间的传播。数据来源的真实性有待查证，无处可查是当下数据新闻中存在的一个弊病。

3.2 数据的分析与呈现

数据新闻不同于新闻数据，数据新闻中的数据是海量的，虽然较于传统的抽样调查。具有相对的“全面性”，但是这么大的数量必然会导致一部分数据不准确。因此，在此数据基础上进行分析也就很难实现精确，具有一定的模糊性。

数据的呈现放大了新闻的议程设置功能，由于数据给受众的冲击力是强大的，再加之可视化的呈现方式，使得数据新闻所报道的数据之间的逻辑关系在受众中得到深化。比如央视的《据说春运》一经播出 ，受众就会对春运形成一种形象的感知，也不容易听得进去关于春运相反面的新闻。这无形中加深和固化了受众对我国春运的认知基膜。由于数据新闻的特性，如果没有媒体利用这一特性对受众进行有目的的议程设置，后果无疑是影响巨大。数据新闻的呈现在这一方面不利于新闻对拟态环境的客观塑造。

3.3 把关人角色的加大

在数据新闻中，数据都是客观存在的，数据的分析都是通过工具实现的，数据新闻就是呈现数据分析得出的数据之间的相关关系，这一切看似对把关人的角色是一种淡化，实际上，数据新闻中的把关人的角色不但没有淡化，反而加大。在选题的过程中就是把关人出现并履行职责的时候，一旦选题出现失误，报道就会引领错误的价值观，导致不良的社会舆论。同时，在数据的筛选、清洗过程中，数据的取舍也是把关人作用的体现，一旦数据取舍有误，失实的信息大量涌入有效信息容量之中，数据的相对关系就会产生相应的偏差，这样会造成数据新闻报道引领错误的社会舆论，对社会的正常发展机制造成危害。

反之，在数据新闻中，把关人只要稍稍对数据进行手脚，就会造成较大的社会影响，一旦别有用心的人承担把关人的角色，后果不堪想象。

4. 对策

对数据新闻生产过程的参与者而言，要加强理论学习，提高新闻职业素养，同时还要学习多种技能，随着数据新闻的发展，仅仅只会采编的记者可能不适应技术性较强的数据新闻生产，因此，记者要不断地提高自己的技能。提高新闻职业素养是新闻从业者在职业生涯永恒不变的要求，在数据新闻领域，凡是参与新闻生产的人员都应提高新闻职业素养，只有具有较高的职业素养的人参与到新闻生产的过程中去，才能更好的履行把关人的角色，才能在数据新闻报到中做到平衡报道，追求数据新闻作为新闻本身的价值和意义。

数据来源开放共享是解决数据来源、拓展数据新闻报道角度的重要途径。数据一旦开放共享，数据的失实和误差将会降到最低，大量的数据也可以为新闻媒体挖掘数据背后不同的相关关系，形成丰富的报道主题。这也可以杜绝某些媒体因独占数据，其他媒体无法核实该媒体基于独占数据的报道是否合理，而形成的“一言堂”现象。虽然在现阶段还不可能实现，但这是数据新闻良好发展的一个希冀。

数据新闻在选择合适的可视化呈现方式时，要充分考虑受众的理解程度。数据新闻尽管具有可视化的直观性，但是对专业领域的报道一定要匹配相应的通俗的揭示文字，否则非专业的受众就会对此信息失去阅读的兴趣，本来直观的呈现优势也不存在。即使是非专业领域的数据新闻报道，也要充分结合故事性来叙述，这样不至于数据报道干涩，没有活力。结合了故事性的数据新闻呈现方式既迎合了受众的理解力，也使得数据新闻有了鲜活的张力。

结语

数据新闻为新闻事业的发展注入了新鲜的血液 ，也促进了媒介融合的时代主题，数据驱动新闻使得新闻的故事性更加客观和直观，对受众而言，也符合当下快阅读的习惯。但是，数据新闻还在不断的摸索中发展，也存在不少的弊端，给学界和业界提出了新的探究课题。媒

（作者单位：河北经贸大学）